

PATENT ABSTRACTS OF JAPAN

(11)Publication number: 10207759 A

(43)Date of publication of application: 07.08.98

(51)Int. Cl G06F 12/00
G06F 12/00
G06F 17/30

(21)Application number: 09011404

(71)Applicant: SHARP CORP

(22)Date of filing: 24.01.97

(72)Inventor: ISHITANI TAKASHI

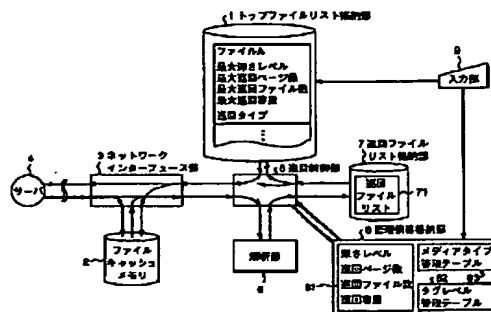
(54)DEVICE FOR AUTOMATICALLY OBTAINING
HYPER TEXT

COPYRIGHT: (C)1998,JPO

(57)Abstract:

PROBLEM TO BE SOLVED: To obtain a page valid to a user in the small number of cache areas.

SOLUTION: A top file and an access condition including the maximum depth level are stored so as to be made correspond to each other in a top file list storage part 1. A network interface part 3 issues a file capturing request to a server 4 according to a request from a cyclic control part 5, and stores received file data in a file cache memory 2, and returns it to the cyclic control part 5. An analyzing part 6 analyzes the file data, and extracts a relevant file name. The cyclic control part 5 calculates the depth level of the relevant file from the analyzing part 6, and stores it so as to be made correspond to the relevant file name in a cyclic list storage part 7, and requests the capture of the file data, and monitors access to a network so as not to be beyond the access condition including the maximum depth level. Thus, the necessary file can be automatically obtained by flexibly limiting an access range according to a cache area.



【特許請求の範囲】

【請求項1】 回線によってネットワーク状に結ばれたデータベースに格納されたハイパーテキストのファイルデータを自動的に取得するハイパーテキスト自動取得装置であって、取得すべきハイパーテキストのトップファイルのファイル名が格納されたトップファイル格納部と、関連ファイルのファイル名が格納される関連ファイルリスト格納部と、上記ネットワーク上を辿る際の最大深さレベルを含む上記ネットワークへのアクセス条件が、上記トップファイル格納部に格納された各ファイル名に対応付けられて格納されたアクセス条件格納部と、受信したファイルデータが格納されるファイルキャッシュメモリと、受けた取得要求に従って上記回線を介して上記ネットワーク上のサーバに接続し、上記接続されたサーバに対して上記トップファイルおよび関連ファイルのファイルデータの取得要求を出す一方、上記サーバから返送されてくるファイルデータを上記ファイルキャッシュメモリに格納すると共に、取得要求先に返送するネットワークインターフェース部と、受けたファイルデータを解析して関連ファイル名を抽出し、抽出した関連ファイル名を返送する解析部と、上記ネットワークインターフェース部、解析部及び関連ファイルリスト格納部を制御して、上記アクセス条件格納部に格納されているアクセス条件に従って、上記トップファイル格納部にファイル名が格納されているトップファイルとこのトップファイルの関連ファイルとを取得するファイル取得制御部を備えたことを特徴とするハイパーテキスト自動取得装置。

【請求項2】 請求項1に記載のハイパーテキスト自動取得装置において、上記ファイルデータに含まれて2つのファイル間の関連状態を表すタグと、このタグが表す関連状態に従って上記ネットワーク上を関連元のファイルから関連先のファイルまで辿る際の深さレベルを表すタグレベルとが、互いに対応付けられて格納されたタグレベル管理テーブルを備えて、上記解析部は、ファイルデータに含まれるタグが表す関連状態に基づいて上記関連ファイル名を抽出し、この抽出した関連ファイル名に上記タグを付加して上記ファイル取得制御部に返送する関連ファイル抽出手段を有し、上記ファイル取得制御部は、上記解析部から送出されたタグに基づいて、上記タグレベル管理テーブルを参照して上記関連ファイルの上記ネットワーク上での深さレベルを算出する深さレベル算出手段と、この算出した深さレベルを対応する関連ファイル名に付加して上記関連ファイルリスト格納部に格納する関連ファイル格納手段と、上記関連ファイルの取得要求を出す際に当該関連フ

ファイルの上記ネットワーク上での深さレベルが上記最大深さレベルを越えるか否かを監視して、越える場合にはファイルデータの取得を中止する深さレベル監視手段を有することを特徴とするハイパーテキスト自動取得装置。

【請求項3】 請求項2に記載のハイパーテキスト自動取得装置において、

上記タグは、2つのファイルが共に同一ページを構成するような関連状態を表すタグと、2つのファイルが異なるページを構成するような関連状態を表すタグを含むことを特徴とするハイパーテキスト自動取得装置。

【請求項4】 請求項2に記載のハイパーテキスト自動取得装置において、

上記タグは、2つのファイルが共に同一ページを構成し、且つ、異なる時間属性を有するような関連状態を表すタグを含んでおり、

上記ファイル取得制御部の深さレベル算出手段は、上記タグが有する時間情報を深さレベルの算出値に反映させるようになっていることを特徴とするハイパーテキスト自動取得装置。

【請求項5】 請求項2に記載のハイパーテキスト自動取得装置において、

上記アクセス条件は、上記トップファイルが含まれるトップページから見て上記ネットワーク上での深さが浅い順にアクセスする第1のアクセス方法と、上記トップページから順次上記深さの深い方にアクセスする第2のアクセス方法とを有する上記ネットワークに対するアクセス方法を含み、

上記ファイル取得制御部の関連ファイル格納手段は、上記アクセス条件格納部に上記第1のアクセス方法が格納されている場合には、関連ファイル名と深さレベルとを上記関連ファイルリスト格納部に追加する毎に上記深さレベルの昇順にソートする一方、上記第2のアクセス方法が格納されている場合には、関連ファイル名と深さレベルとを上記関連ファイルリスト格納部の最後尾に追加するようになっており、

上記ファイル取得制御部は、上記アクセス条件格納部に上記第1のアクセス方法が格納されている場合には、関連ファイル取得要求時に上記関連ファイルリスト格納部の先頭から順に関連ファイル名と深さレベルとを読み出す一方、上記第2のアクセス方法が格納されている場合には、上記関連ファイルリスト格納部の最後尾から順に読み出すようになっていることを特徴とするハイパーテキスト自動取得装置。

【請求項6】 請求項1に記載のハイパーテキスト自動取得装置において、

ファイルデータのメディアタイプとこのメディアタイプに属するファイルデータを取得するか否かを表す取得情報とが対応付けられて格納された取得情報格納部を備え

上記ネットワークインターフェース部は、上記ファイル取得制御部からの指示に従って、上記ファイルキャッシュメモリに格納されたファイルデータを削除するファイル削除手段を有し、

上記ファイル取得制御部は、受け取ったファイルデータのメディアタイプを解釈するメディアタイプ解釈手段と、上記メディアタイプの解釈結果に基づいて上記取得情報格納部の取得情報を参照して取得しないと判定したファイルのファイルデータの削除を上記ネットワークインターフェース部に指示する削除指示手段を有することを特徴とするハイパーテキスト自動取得装置。

【請求項7】 請求項1に記載のハイパーテキスト自動取得装置において、

上記ネットワークインターフェース部は、上記ファイル取得制御部からの指示に従って、上記ファイルキャッシュメモリに格納されたファイルデータを削除するファイル削除手段を有し、

上記アクセス条件は、各トップファイル毎に定められた最大取得容量を含んでおり、

上記ファイル取得制御部は、上記関連ファイルを受け取る毎に、この受け取った関連ファイルを加味した取得ファイルの総容量が上記最大取得容量を越えるか否かを監視し、越える場合には、最後に受け取ったファイルデータの削除を上記ネットワークインターフェース部に指示するファイル容量監視手段を有することを特徴とするハイパーテキスト自動取得装置。

【請求項8】 請求項1、請求項2および請求項6の何れか一つに記載のハイパーテキスト自動取得装置において、

入力部を備えて、

上記トップファイル格納部に格納されるファイル名、上記アクセス条件格納部に格納されるアクセス条件、上記タグレベル管理テーブルの要素、あるいは、上記取得情報格納部の要素は、上記入力部から入力・更新可能になっていることを特徴とする。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、WWW(World Wide Web)等のハイパーテキストのデータをネットワークを経由してサーバから取得するハイパーテキスト自動取得装置に関する。

【0002】

【従来の技術】 従来より、データ取得方式として、センターからの要求に対してコマンド形式で応答するセンター方式がある。このセンター方式においては、予め、センターから送出されてくる種々のコマンド要求とこのコマンド要求に呼応して送信(実行)すべきコマンドとを対応付けてクライアント装置に記憶しておく。こうすることによって、目的とするコマンドを自動的に送出して、目的とするデータを取得できるのである。このようなセ

ンター方式は、パソコン通信ターミナルソフト等に実装されている。

【0003】 これとは別に、物理的な位置によらずにネットワーク上に分散されたWWW等のハイパーテキストのデータを取得する場合には、サーバからクライアント装置に対して特に要求は出されない。したがって、クライアント装置から能動的に必要な要求を出す必要がある。この場合、上記クライアント装置にはユーザインターフェース・ツール「ブラウザ」が搭載されている。そして、ネットワーク上における目的のサーバ名とファイル名とを指定すると、上記ブラウザによって、指定されたファイルとそのファイルをメインとしたページを構成する他のファイルとが自動的に取得されて、ビジュアルに組み合わせて当該ページが表示される。

【0004】 ここで、上記指定のファイルをメインとするページが他のページと関連する場合に、互いに関連するページ間を移動して表示させるためには、1つのページから関連するページへのリンク(図9に示すリンク「anchor」等)等をユーザが1つずつ指定して順に表示する必要がある。また、扱うファイル群は、ネットワーク状のあらゆる場所に散在する。したがって、上記WWW等のハイパーテキストのデータを取得する場合には、実際に要求を出してから取得された総てのページの表示が終了するまではかなりの時間を要するのである。ところが、この場合の通信は電話回線を使用するために通信に掛かる単位コストが大きく、そのために短時間に指定のファイルに関連する全ページの取得/表示が終了する機能が望まれている。

【0005】 そこで、このような要求を満たす機能として自動巡回機能がある。この自動巡回機能を有するクライアント装置では、取得した指定ファイルが存在するページのデータ(指定ページ)と関連ページのデータとを記憶するようになっており、記憶後は通信を遮断することができるのである。したがって、その後、ユーザは時間を気にすることなく、指定ページや関連ページとを表示できるのである。

【0006】 すなわち、上記自動巡回機能では、以下のことを行うのである。

- ・指定ページから、ハイパーテキストのネットワーク構造によってリンクされた関連ページを取得する。
- ・指定されたページ取得時間内にページを取得する。
- ・ページを取得した後、即座にビューワと連動して取得した目的ページと関連ページとを表示する。

【0007】 以下、上記自動巡回機能を具体的な例を上げて説明する。

(1) ハイパーメディア文書通信装置(特開平8-87526号公報)

このハイパーメディア文書通信装置では、ハイパーテキストを構成する関連するシーンを結ぶリンクが過去に何回選択されたかを表す履歴情報が保持/管理されてい

る。そして、事前転送シーン判断部によって、プレイ要求のあった表示シーンS0からリンクが張られている一次、二次、…i次のシーンの各々が事前に転送すべきシーンであるか否かを判断するための関連度値が、該当するリンクの履歴情報と表示シーンS0から目的シーンSiまでの距離とに基づいて算出される。そして、算出された関連度値が閾値より大きいシーンが事前転送候補として選出される。こうして選出されたシーンをハイパー文書表示端末のメモリに事前転送することによって、ユーザがボタンを操作して関連シーンをアクセスした場合の読み出し速度を向上できるのである。

【0008】(2) ソフトウェア「フリーローダ」(Free Loader, Inc. URL: <http://www.freeloader.com/>) この「フリーローダ」はパソコン上で起動するソフトウェアであり、WWW上の指定ページを自動的に取得して記憶装置に保存するソフトウェアである。このソフトウェアは、指定した時間に自動的に起動させることができる。また、このソフトウェアでは、予め幾つものページが分類されており、ユーザは分類項目を指定するだけで指定分類項目に該当するページを取得できるようになっている。

【0009】(3) ソフトウェア「波乗野郎」(B. U. G., Inc. URL: <http://www.bug.co.jp/nami-nori/index.html>)

この「波乗野郎」もパソコン上で起動するソフトウェアであり、次のような特徴を有する。

- ・指定時間(開始時間と終了時間)に指定のページを自動的に取得する。
- ・取得可能なページへのリンク数を予め指定できる。
- ・ハイパーテキストのリンクを辿る際に、指定されたサーバに関するリンクのみを辿るようにできる。

【0010】(4) ソフトウェア「インターネットマネージャ」(日本電機URL: <http://www.nec.co.jp/japanese/product/personal/product/inet/iam/iam-pl.html>) この「インターネットマネージャ」もパソコン上で起動するソフトウェアであり、次のような特徴を有する。

- ・指定時間に自動的に指定ページを取得できる。
- ・利用時間および料金の制限を設定できる。
- ・複数の接続先(電話番号)を優先順位を付けて登録できる。

【0011】

【発明が解決しようとする課題】ところで、ハイパーテキストの構造は、図9に示すように、あるページが関連ページとリンクされてネットワーク構造を呈している。したがって、自動巡回機能によって取得される指定ページと関連ページとには実際にユーザが欲しないデータもかなり多く含まれて膨大な数となり、その記憶装置も膨大な記憶容量が必要である。特に、携帯型の装置の場合には、外部記憶装置の追加が困難であるために、取得したページを蓄えておくためのキャッシュエリアも多くは

とれない。

【0012】上述した自動巡回機能を有するハイパーメディア文書通信装置や自動巡回機能を有するソフトウェアにおいては、関連度の高いページを取得したり、指定された項目のページのみを取得したり、指定されたリンク数内でページを取得したりして、取得ページ数のある程度絞るようにはしている。ところが、この程度の絞りでは絞られる取得ページ数にも限度がある。また、時間や利用料金で取得ページ数に制限を加えた場合には、必ずしもユーザにとって有効なページが取得されるとは限らない。すなわち、従来の自動巡回機能を有するハイパーメディア文書通信装置や自動巡回機能を有するソフトウェアでは、携帯型装置のような少ないキャッシュエリアでユーザにとって有効なページを取得するためには問題がある。

【0013】そこで、この発明の目的は、少ないキャッシュエリアでユーザに有効なページを効率よく自動的に取得できるハイパーテキストの自動取得装置を提供することにある。

【0014】

【課題を解決するための手段】上記目的を達成するため、請求項1に係る発明は、回線によってネットワーク状に結ばれたデータベースに格納されたハイパーテキストのファイルデータを自動的に取得するハイパーテキスト自動取得装置であって、取得すべきハイパーテキストのトップファイルのファイル名が格納されたトップファイル格納部と、関連ファイルのファイル名が格納される関連ファイルリスト格納部と、上記ネットワーク上を辿る際の最大深さレベルを含む上記ネットワークへのアクセス条件が、上記トップファイル格納部に格納された各ファイル名に対応付けられて格納されたアクセス条件格納部と、受信したファイルデータが格納されるファイルキャッシュメモリと、受けた取得要求に従って上記回線を介して上記ネットワーク上のサーバに接続し、上記接続されたサーバに対して上記トップファイルおよび関連ファイルのファイルデータの取得要求を出す一方、上記サーバから返送されてくるファイルデータを上記ファイルキャッシュメモリに格納すると共に、取得要求先に返送するネットワークインターフェース部と、受けたファイルデータを解析して関連ファイル名を抽出し、抽出した関連ファイル名を返送する解析部と、上記ネットワークインターフェース部、解析部及び関連ファイルリスト格納部を制御して、上記アクセス条件格納部に格納されているアクセス条件に従って、上記トップファイル格納部にファイル名が格納されているトップファイルとこのトップファイルの関連ファイルとを取得するファイル取得制御部を備えたことを特徴としている。

【0015】上記構成において、ファイル取得制御部によって、トップファイル格納部に格納されたトップファイルの取得要求がネットワークインターフェース部に出

される。そうすると、上記ネットワークインターフェース部は、ファイルキャッシュメモリに上記取得要求のあったファイルが登録されていない場合には、受けた取得要求に従ってネットワーク上のサーバに接続して上記トップファイルの取得要求を出す。そして、上記サーバから返送されてきたファイルが上記ファイルキャッシュメモリに格納されると共に、ファイル取得制御部に返送される。

【0016】次に、上記ファイル取得制御部によって、受け取ったトップファイルのファイルデータが解析部に送出される。そして、上記解析部によって、受け取ったトップファイルのファイルデータが解析されて、当該トップファイルの関連ファイル名が抽出されて上記ファイル取得制御部に返送される。そうすると、上記ファイル取得制御部によって、上記抽出された関連ファイルのファイル名が関連ファイルリスト格納部に格納される。以後、上記ファイル取得制御部によって、アクセス条件格納部に格納されているアクセス条件に従って、上記トップファイルの場合と同様に、関連ファイル格納部に格納された関連ファイルの取得要求が上記ネットワークインターフェース部に出され、取得された関連ファイルは上記ファイルキャッシュメモリに格納される。

【0017】こうして、上記トップファイル毎に設定されたアクセス条件に従ってハイパーテキストのネットワーク上を辿って関連ファイルを取得するので、各トップファイル毎に柔軟且つ的確に制限された範囲内でのアクセスが行われる。特に、上記アクセス条件には最大深さレベルが含まれているので、上記ネットワーク上を辿る際の深さレベルが上記トップファイル毎に制限されて、ファイルキャッシュメモリの容量内で必要なファイルが的確に自動的に取得されるのである。

【0018】また、請求項2に係る発明は、請求項1に係る発明のハイパーテキスト自動取得装置において、ファイルデータに含まれて2つのファイル間の関連状態を表すタグと、このタグが表す関連状態に従って上記ネットワーク上を関連元のファイルから関連先のファイルまで辿る際の深さレベルを表すタグレベルとが、互いに対応付けられて格納されたタグレベル管理テーブルを備えて、上記解析部は、ファイルデータに含まれるタグが表す関連状態に基づいて上記関連ファイル名を抽出し、この抽出した関連ファイル名に上記タグを付加して上記ファイル取得制御部に返送する関連ファイル抽出手段を有し、上記ファイル取得制御部は、上記解析部から送出されたタグに基づいて、上記タグレベル管理テーブルを参照して上記関連ファイルの上記ネットワーク上での深さレベルを算出する深さレベル算出手段と、この算出した深さレベルを対応する関連ファイル名に付加して上記関連ファイルリスト格納部に格納する関連ファイル格納手段と、上記関連ファイルの取得要求を出す際に当該関連ファイルの上記ネットワーク上での深さレベルが上記最

大深さレベルを越えるか否かを監視して、越える場合にはファイルデータの取得を中止する深さレベル監視手段を有することを特徴としている。

【0019】上記構成によれば、関連ファイル名に深さレベルを付けて関連ファイルリスト格納部に登録されている。したがって、後にファイル取得制御部によって実行される関連ファイルの取得要求時に、深さレベル監視手段によって、取得しようとする関連ファイルのハイパーテキストのネットワーク上での深さレベルが上記アクセス条件の1つの最大深さレベルを越えないように監視されて、上記ネットワーク上におけるアクセス範囲が上記ネットワークの深さレベルで制限される。

【0020】また、請求項3に係る発明は、請求項2に係る発明のハイパーテキスト自動取得装置において、上記タグは、2つのファイルが共に同一ページを構成するような関連状態を表すタグと、2つのファイルが異なるページを構成するような関連状態を表すタグを含むことを特徴としている。

【0021】上記構成によれば、2つのファイルが共に同一ページを構成するような関連状態、および、2つのファイルが異なるページを構成するような関連状態が区別されて、夫々のタグレベルが設定される。

【0022】また、請求項4に係る発明は、請求項2に係る発明のハイパーテキスト自動取得装置において、上記タグは、2つのファイルが共に同一ページを構成し、且つ、異なる時間属性を有するような関連状態を表すタグを含んでおり、上記ファイル取得制御部の深さレベル算出手段は、上記タグが有する時間情報を深さレベルの算出値に反映させるようになっていることを特徴としている。

【0023】上記構成によれば、2つのファイルが同一ページを構成し、且つ、異なる時間属性を有するような関連状態を示す関連先のファイルの深さレベルが、当該関連状態を表すタグが有する時間情報を反映して算出される。したがって、例えば、表示に時間が掛かるような関連ファイルの深さレベルの値を大きくすることによって、取得の優先度が低められる。

【0024】また、請求項5に係る発明は、請求項2に係る発明のハイパーテキスト自動取得装置において、上記アクセス条件は、上記トップファイルが含まれるトップページから見て上記ネットワーク上での深さが浅い順にアクセスする第1のアクセス方法と、上記トップページから順次上記深さの深い方にアクセスする第2のアクセス方法とを有する上記ネットワークに対するアクセス方法を含み、上記ファイル取得制御部の関連ファイル格納手段は、上記アクセス条件格納部に上記第1のアクセス方法が格納されている場合には、関連ファイル名と深さレベルとを上記関連ファイルリスト格納部に追加する毎に上記深さレベルの昇順にソートする一方、上記第2のアクセス方法が格納されている場合には、関連ファイ

ル名と深さレベルとを上記関連ファイルリスト格納部の最後尾に追加するようになっており、上記ファイル取得制御部は、上記アクセス条件格納部に上記第1のアクセス方法が格納されている場合には、関連ファイル取得要素時に上記関連ファイルリスト格納部の先頭から順に関連ファイル名と深さレベルとを読み出す一方、上記第2のアクセス方法が格納されている場合には、上記関連ファイルリスト格納部の最後尾から順に読み出すようになっていることを特徴としている。

【0025】上記構成によれば、上記アクセス条件格納部に第1のアクセス方法が格納されている場合には、上記トップファイルが含まれるトップページから見て上記ネットワーク上での深さが浅い順にアクセスが行われて、広く浅く関連ファイルが取得される。これに対して、第2のアクセス方法が格納されている場合には、上記トップページから順次上記深さの深い方にアクセスが行われて、狭く深く関連ファイルが取得される。こうして、ハイパーテキストのネットワークからの広く浅い関連ファイルの取得と狭く深い関連ファイルの取得とが、指定アクセス方法に従って選択的に行われて、限られたキャッシュエリア内で有効な関連ファイルが取得される。

【0026】また、請求項6に係る発明は、請求項1に記載のハイパーテキスト自動取得装置において、ファイルデータのメディアタイプとこのメディアタイプに属するファイルデータを取得する可否かを表す取得情報とが対応付けられて格納された取得情報格納部を備えると共に、上記ネットワークインターフェース部は、上記ファイル取得制御部からの指示に従って上記ファイルキャッシュメモリに格納されたファイルデータを削除するファイル削除手段を有し、上記ファイル取得制御部は、受け取ったファイルデータのメディアタイプを解釈するメディアタイプ解釈手段と、上記メディアタイプの解釈結果に基づいて上記取得情報格納部の取得情報を参照して取得しないと判定したファイルのファイルデータの削除を上記ネットワークインターフェース部に指示する削除指示手段を有することを特徴としている。

【0027】上記構成によれば、取得情報格納部に予め設定された取得情報に応じて取得されるファイルデータのメディアタイプが限定されて、ユーザが必要とするメディアタイプのファイルのみが取得される。

【0028】また、請求項7に係る発明は、請求項1に係る発明のハイパーテキスト自動取得装置において、上記ネットワークインターフェース部は、上記ファイル取得制御部からの指示に従って、上記ファイルキャッシュメモリに格納されたファイルデータを削除するファイル削除手段を有し、上記アクセス条件は、各トップファイル毎に定められた最大取得容量を含んでおり、上記ファイル取得制御部は、上記関連ファイルを受け取る毎に、この受け取った関連ファイルを加味した取得ファイルの

総容量が上記最大取得容量を越えるか否かを監視し、越える場合には最後に受け取ったファイルデータの削除を上記ネットワークインターフェース部に指示するファイル容量監視手段を有することを特徴としている。

【0029】上記構成によれば、特に携帯型の装置であってファイルキャッシュメモリの容量が大きい場合でも、上記最大取得容量を最適に設定することによって、上記容量内で取得し得る最大数のファイルデータが取得される。

10 【0030】また、請求項8に係る発明は、請求項1、請求項2および請求項6の何れか一つに係る発明のハイパーテキスト自動取得装置において、入力部を備えて、上記トップファイル格納部に格納されるファイル名、上記アクセス条件格納部に格納されるアクセス条件、上記タグレベル管理テーブルの要素、あるいは、上記取得情報格納部の要素は、上記入力部から入力・更新可能になっていることを特徴としている。

20 【0031】上記構成によれば、入力部からトップファイル格納部、アクセス条件格納部、タグレベル管理テーブルあるいは取得情報格納部の内容がユーザによって更新設定されて、ハイパーテキストのネットワーク上へのアクセス範囲が各トップファイル毎に柔軟且つ的確に制限される。

【0032】

【発明の実施の形態】以下、この発明を図示の実施の形態により詳細に説明する。図1は、本実施の形態のハイパーテキスト自動取得装置における機能ブロック図である。また、図2は、図1の機能を実現するためのハードウェアブロック図である。

30 【0033】図1において、トップファイルリスト格納部1には、ユーザによって巡回することが指示されたファイル名のリストが格納される。また、トップファイルリスト格納部1には、上記ファイル名に対応付けて最大深さレベル、最大巡回ページ数、最大巡回ファイル数、最大巡回容量、巡回タイプ等の上記ハイパーテキストのネットワークをアクセスする際のアクセス条件が格納される。ファイルキャッシュメモリ2には、取得されたファイルデータが保存される。こうして、ハイパーテキストのネットワーク上に分散しているファイルのデータをファイルキャッシュメモリ2に保存しておくことによって、後に同じファイルを取得しようとする際には上記ネットワークにアクセスせずに、ファイルキャッシュメモリ2から直接データを読み出すことによって、同じファイルに対する2回目以降の取得速度を高速可能にするのである。

50 【0034】ネットワークインターフェース部3は、ネットワーク上のサーバ4と巡回制御部5との橋渡しを行うものである。このネットワークインターフェース部3は、巡回制御部5からファイルの取得要求を受け取ると、ファイルキャッシュメモリ2内に上記取得要求が出

されたファイルのデータが保存されているか否かを判別する。そして、存在する場合には、ファイルキャッシュメモリ2から取得要求があったファイルのデータを読み出して巡回制御部5に返送する。これに対して、存在しない場合には、サーバ4に対してファイル要求を出す。そして、取得要求したファイルのデータがサーバ4から返送されてくると、そのデータをファイルキャッシュメモリ2に追加しつつ巡回制御部5に返送する。

【0035】解析部6は、文書ファイルの構造を解析して、取得された文書ファイル内に記述されたタグや関連ファイル名を抽出する。ここで、上記関連ファイルとは、取得された文書ファイル内に埋め込むべきイメージファイル(IMGタグが付されている)、取得された文書ファイルに関連する別文書ファイル(ANCHORタグが付されている)、時間属性を有して同一文書ファイルにおける異なる時間に在る文書ファイル(METAタグが付されている)等であり、上述のような種々のタグを付けることによって、所謂「img」リンクや「anchor」リンクや「meta」リンクで連結された(関連付けられた)ファイルである。また、上述のごとく、上記タグは2つのファイル間の関連状態を表しており、夫々のタグが付されたファイルには、そのタグが表す関連状態に適合した関連ファイルが関連付けられている。すなわち、解析部6は、取得文書ファイルに記述されたタグとこのタグが付されているファイルとを対応付けることによって、関連ファイル名を抽出するのである。

【0036】巡回ファイルリスト格納部7には、図3に示すような巡回ファイルリスト71が格納される。この巡回ファイルリスト71には、図3に示すように、上記解析部6によって抽出された関連ファイル名とその関連ファイルが存在するハイパーテキストのネットワーク上での深さレベルとが対応付けられて登録される。

【0037】管理情報格納部8には、ハイパーテキストのネットワーク上において、現在巡回制御部5が処理中のファイルに関する深さレベル(現在の巡回深さレベル)、巡回ページ数、巡回ファイル数、巡回容量等の制御パラメータ81、図4に示すようなタグレベル管理テーブル82、および、図5に示すようなメディアタイプ管理テーブル83が格納される。

【0038】ここで、上記管理情報中の制御パラメータ「深さレベル」とは、処理中のファイルに関するハイパーテキストのネットワーク上での深さを表し、トップファイルリスト格納部1から読み出されたファイル(トップファイル)に関するファイルデータの処理を行う場合の深さレベルを「1」とする。そして、各関連ファイルの深さレベルを、その関連ファイルの関連状態に応じて図10に示すように設定する。例えば、上記トップページとは別のページを構成する関連ファイルの場合には、深さレベルを「2」とするのである。また、上記制御パラメータ「巡回ページ数」は現在まで幾つのページに属

するファイルを処理したかを表し、1ページを構成する複数のファイルに関する処理数は「1」としてカウントする。また、制御パラメータ「巡回ファイル数」は、実際に巡回して処理したファイル数を表す。また、制御パラメータ「巡回容量」は、巡回して取得したファイルに関するデータの総容量を表す。これらの制御パラメータ81を用いることによって、上述のごとく、外部から与えられてトップファイルリスト格納部1に登録される上記制御パラメータの最大値(アクセス条件)に従って、巡回制御部5の巡回動作が容易に且つ最適に制御されるのである。

【0039】また、上記タグレベル管理テーブル82は、上記タグによって関連付けられた関連ファイルへのアクセスを管理する。例えば、関連ファイルへアクセスする際の深さレベルを例に取ると、あるタグによって関連付けられた関連ファイルの取得処理に移行する場合に、現在の巡回深さレベルの値をどれだけ上げるべきかを管理するのである。

【0040】したがって、本実施の形態によれば、上記タグレベル管理テーブル82の内容を更新するだけで、新たなタグが新設されても、タグが表す関連状態が変更されても、柔軟に対応できるのである。また、「タグレベル」値として小数点以下の値を設定可能にすることによって、通常の深さレベルの単位「1」の中間階層を設定できる。また、上記タグレベル値として深さレベルの単位「1」の中間階層を設定することによって、本来ならば上記タグレベル値として「0」を設定すべき所を無限連鎖を避けるために小数点以下の値を設定することが可能となるのである。ここで、上記「無限連鎖」とは、ハイパーテキストのネットワーク上を順次辿ってファイルをアクセスしている際に、(例えば、タグレベル=0が設定されているために)上記ネットワーク上での深さレベルが変わらなくなり、アクセスが上記ネットワーク上を無限に連鎖して行く状態のことを言う。

【0041】また、上記メディアタイプ管理テーブル83には、ファイルデータのメディアタイプとそのメディアタイプに属するファイルデータの取得の可否を表す取得フラグとが、互いに対応付けられて登録されている。このメディアタイプ管理テーブル83によって、受信したファイルデータを取得するか否かを管理するのである。

【0042】上記巡回制御部5は、上記トップファイルリスト格納部1に格納されたトップファイルのリストに従って、次のようにして、ユーザによって指定されたファイルのデータを取得する。

【0043】すなわち、上記トップファイルリスト格納部1あるいは巡回ファイルリスト格納部7から巡回すべきファイルのファイル名を読み出して、ネットワークインターフェース部3に当該ファイルの取得要求を出す。そして、ネットワークインターフェース部3からデータ

が返送されてくると、そのデータタイプがハイパーテキストタイプであるか否かを判断する。そして、ハイパーテキストタイプでなければ受け取ったデータを破棄し、ハイパーテキストタイプのデータである場合には解析部6にデータを転送して関連ファイル名を抽出させる。そして、解析部6から受け取った関連ファイル名を巡回ファイルリスト格納部7に追加・登録するのである。その場合のファイルデータ取得処理は、上記トップファイルリスト格納部1に格納された上記アクセス条件および管理情報格納部8に格納された管理情報に従って行われる。

【0044】また、上記巡回制御部5は、上記巡回ファイルリスト格納部7に関連ファイル名と共に格納する深さレベルを、管理情報格納部8に格納されたタグレベル管理テーブル82のタグレベルに基づいて次式によって算出する。深さレベル＝現在の巡回深さレベル＋対応するタグの「タグレベル」

【0045】また、上記巡回制御部5は、トップファイルリスト格納部1に上記アクセス条件の1つとして格納された次の2つの巡回タイプの何れかに従って、巡回ファイルリスト格納部7に登録された関連ファイルを取得する。

(1) タイプA

タイプAでは、上記巡回ファイルリスト71の内容を、新たな関連ファイル名を追加する毎に各関連ファイル名に対応付けられた深さレベルの昇順にソートしておく。そして、ハイパーテキストのネットワークにアクセスする際には先頭の関連ファイル(つまり、深さレベルが最も小さい関連ファイル)の順にアクセスする。そして、アクセスした関連ファイル名は、巡回ファイル管理リストから削除しておく。したがって、このタイプAは、トップファイルリスト格納部1による指定に基づいて取得したファイル(トップファイル)をメインとしたページ(トップページ)から見て深さレベルの浅い順に横方向にアクセスする横方向巡回方式と言える。

【0046】(2) タイプB

タイプBでは、上記巡回ファイルリスト71に新たな関連ファイル名を追加する場合には最後尾に追加する。そして、上記ネットワークにアクセスする際には最後尾の関連ファイルから順にアクセスする。そして、アクセスした関連ファイル名は巡回ファイル管理リストから削除しておく。したがって、このタイプBは、トップページからリンクを伝って順次深さレベルの深い方にアクセスする縦方向巡回方式と言える。

【0047】こうして、巡回ファイルリスト格納部7に登録された関連ファイルが総て削除されると、トップファイルリスト格納部1から読み出されたファイル名に関するトップページと関連ページの取得を終了するのである。

【0048】入力部9からは、上記トップファイルリス

ト格納部1に登録されるトップファイル名や上記アクセス条件、管理情報格納部8に格納されるタグレベル管理テーブル82およびメディアタイプ管理テーブル83の要素が、マニュアル設定される。このように、上記アクセス条件やタグレベル管理テーブル82およびメディアタイプ管理テーブル83の要素を外部から指定することによって、ハイパーテキストのネットワーク上におけるアクセスすべき範囲を柔軟に制限・変更できるのである。

10 【0049】本実施の形態におけるハイパーテキスト自動取得装置におけるハードウェア構成は、図2に示すようになっている。ネットワーク装置11は、本ハイパーテキスト自動取得装置を通信/ネットワーク12に接続する。ネットワーク装置11を介して通信/ネットワーク12から取得されたファイルデータは外部記憶装置13に格納される。外部記憶装置14には、指定ファイルへのアクセス方法、上記アクセス条件、取得すべき関連ファイル名および管理情報等が格納される。また、RAM(ランダム・アクセス・メモリ)で構成されたメモリ15は作業用ワークメモリとして使用される。

20 【0050】CPU16は、ROM(リード・オンリ・メモリ)で構成されたプログラムメモリ17に格納された制御プログラムに従って、上記ネットワーク装置11、外部記憶装置14およびメモリ15等を制御して、上記トップファイルおよび関連ファイルを取得する。入力装置18は、上記CPU16の動作を最適に制御するためのアクセス条件や各種パラメータや指令を入力する。

30 【0051】すなわち、上記ネットワーク装置11によって図1におけるネットワークインターフェース部3を構成し、外部記憶装置13でファイルキャッシュメモリ2を構成し、外部記憶装置14でトップファイルリスト格納部1、巡回ファイルリスト格納部7および管理情報格納部8を構成し、CPU16で巡回制御部5および解析部6を構成し、入力装置18で入力部9を構成しているのである。

【0052】以下、上記巡回制御部5、ネットワークインターフェース部3および解析部6によって行われる巡回制御処理動作について、詳細に説明する。以下においては、ハイパーテキストとして、HTML(Hyper Text Markup Language)で記述された文書を例にして説明する。HTMLで記述された文書では、物理的なファイルがネットワーク上のどこに存在するかを一意に表現するために、URL(Uniform Resource Location)という表記法でファイル名を表記する。このURLは、次のようなフォーマットを有する表記法である。

proto://server/path/...

ここで、proto:ネットワーク上をアクセスためのプロトコル名(通常、http(HyperTextTransfarProtocol)が用いられる)

50 server:指定のサーバマシン(ホスト)

path: 指定サーバマシン内でのファイルの位置

【0053】また、上記HTMLにおいては、文書ファイル中における他のファイルの参照箇所、下記のようなタグを記述して、上述したような様々な関連状態の意味内容を表すことができる。

…… : ANCHORタグ

【0054】いま、取得しようとするハイパーテキストの構造が図9に示すようになっている場合を例に、図1を用いて、図6～図8のフローチャートに従って巡回制御処理動作を具体的に説明する。

【0055】ステップS1で、上記巡回制御部5によって、上記管理情報格納部8に格納されている制御パラメータ「深さレベル」、「巡回ページ数」、「巡回ファイル数」、「巡回容量」が、次のように初期化される。

深さレベルCurrentLevel = 0

巡回ページ数CurrentPageNum = 0

巡回ファイル数CurrentFileNum = 0

巡回容量CurrentVolume = 0

また、上記制御パラメータ81の最大値が、デフォルト値として上記作業ワークメモリに、例えば次のように自動的に初期設定される。また、次のように管理情報「巡回タイプ」が初期設定される。

最大深さレベルMaxLevel = 3

最大巡回ページ数MaxPageNum = 100

最大巡回ファイル数MaxFileNum = 500

最大巡回容量MaxVolume = 100(kバイト)

巡回タイプ=タイプA

【0056】ステップS2で、上記巡回制御部5によって、トップファイルリスト格納部1から未処理のファイル名とそのファイル名に対応付けられているアクセス条件が読み出される。ここで、読み出されたファイル名を、例えば「ファイルA」とする。尚、このファイル名「ファイルA」は、実際にはURLで記述されている。また、読み出されたファイル名「ファイルA」に、上記アクセス条件「最大深さレベル」、「最大巡回ページ数」、「最大巡回ファイル数」、「最大巡回容量」、「巡回タイプ」が対応付けられている場合には、その対応付けられているアクセス条件で上記デフォルト値および管理情報「巡回タイプ」の内容が更新される。

【0057】ステップS3で、上記巡回制御部5によって、上記管理情報格納部8に格納されている制御パラメータ「深さレベルCurrentLevel」の内容が、「1」に更新される。

CurrentLevel = 1

ステップS4で、上記巡回制御部5によって、ネットワークインターフェース部3に対して、「ファイルA」の取得が要求される。ステップS5で、上記ネットワークインターフェース部3によって、ファイルキャッシュメモリ2が検索されて、ファイルキャッシュメモリ2内に、上記ステップS4で取得が要求された「ファイル

A」、あるいは、後にステップS35で取得が要求される関連ファイルが在るか否かが判別される。その結果、あればステップS6に進み、無ければステップS7に進む。ステップS6で、上記ネットワークインターフェース部3によって、ファイルキャッシュメモリ2から「ファイルA」あるいは「関連ファイル」のファイルデータが読み出されて巡回制御部5に返送される。そうした後ステップS10に進む。

【0058】ステップS7で、上記ネットワークインターフェース部3によって、ファイル名に記述された指定のサーバ4に対して、「ファイルA」あるいは「関連ファイル」のファイル要求が出される。ステップS8で、上記ネットワークインターフェース部3によって、指定のサーバ4からデータが返送されて来たか否かがチェックされる。そして、データの返送があれば、ステップS9に進む。ステップS9で、上記ネットワークインターフェース部3によって、受信したファイルデータのファイルキャッシュメモリ2への格納が開始される。それと同時に、巡回制御部5に返送される。

【0059】ステップS10で、上記巡回制御部5によって、管理情報格納部8に格納されている制御パラメータ「巡回容量CurrentVolume」と上記受け取ったファイルデータの容量とに基づいて、当該ファイルデータを取得した場合に取得ファイル容量が最大巡回容量MaxVolume(=100(kバイト))を越えてしまうか否かが、次式によって判別される。

CurrentVolume+受け取ったファイルデータの容量 < 最大巡回容量MaxVolume

その結果、越えない場合にはステップS12に進む一方、越える場合にはステップS11に進む。ステップS11で、上記巡回制御部5によって、以下のような中断処理が実行される。すなわち、上記ステップS9においてファイルキャッシュメモリ2への格納が開始されたファイルデータ、あるいは、上記ステップS6においてファイルキャッシュメモリ2から読み出されたファイルデータの削除要求がネットワークインターフェース部3に出される。そうすると、この時点までにファイルキャッシュメモリ2に格納された上記受け取ったファイルデータが削除される。そうした後、ステップS24に進む。

【0060】ステップS12で、上記巡回制御部5によって、受け取ったファイルデータのメディアタイプが、管理情報格納部8に格納されたメディアタイプ管理テーブル83を参照してチェックされる。ステップS13で、上記巡回制御部5によって、上記ステップS12におけるチェックの結果に従って、メディアタイプがハイパーテキストタイプであるか否か、つまり、HTMLデータ(メディアタイプが「text/html」)であるか否かが判別される。その結果、HTMLデータであればステップS16に進み、そうでなければステップS14に進む。

【0061】ステップS14で、上記巡回制御部5によっ

て、上記ステップS12におけるチェックの結果を参照して、上記受け取ったファイルデータは取得すべきファイルデータであるか否かが判別される。その結果、取得すべきファイルデータであればステップS15に進み、そうでなければ上記ステップS11に進んで上記中断処理が実行される。ステップS15で、上記巡回制御部5によって、上記受け取ったファイルデータを最後まで読み捨てる読み捨て処理が行われる。その結果、上記ステップS9においてファイルキャッシュメモリ2への格納が開始されたファイルデータが最後まで格納されて取得される。あるいは、上記ステップS6においてファイルキャッシュメモリ2から読み出されたファイルデータが取得される。そうした後、ステップS22に進む。

【0062】ステップS16で、上記巡回制御部5によって、上記ステップS13においてHTMLデータであると判定されたファイルデータが、解析部6に送出される。ステップS17で、上記解析部6によって、以下のようなHTML解析処理が行われる。すなわち、HTMLデータに記述されているタグを読み出し、このタグに基づいて当該HTMLデータのファイルをメインとするページの構成ファイルとの関連状態が得られる。ステップS18で、上記解析部6によって、上記ステップS17において得られた関連状態に基づいて当該HTMLデータの関連ファイル名が抽出され、この抽出された関連ファイル名が上記読み出したタグと共に巡回制御部5に返送される。このステップによって、当該HTMLデータがトップファイルのファイルデータである場合には、関連ファイルとして当該トップファイルの子ファイルが抽出される。また、当該HTMLデータが関連ファイルのファイルデータである場合には、関連ファイルとして上記トップファイルの孫ファイル(ひ孫ファイル、…)が抽出される。

【0063】ステップS19で、上記巡回制御部5によって、受け取ったタグに基づいて、関連ファイルの深さレベルが、上記管理情報格納部8に格納された制御パラメータ「深さレベルCurrentLevel」、および、タグレベル管理テーブル82から得られる「タグレベルTAG_LEVEL」に基づいて次式によって算出される。

$FileLevel = CurrentLevel + TAG_LEVEL$

【0064】ここで、当該関連ファイルのタグは「METAタグ」であって、当該関連ファイルが関連元のファイルと同一ページを構成して、且つ、異なる時間属性を有する場合には、上記解析部6によって、タグから抽出された時間情報「Time」を用いて、次式によって関連ファイルの深さレベルFileLevelを算出する。

$FileLevel = CurrentLevel + TAG_LEVEL \times TimeLevel$

尚、上記「TimeLevel」は、次式によって算出する。

if (Time ≤ StandardTime) then

TimeLevel = 1

else

TimeLevel = Time / StandardTime

すなわち、上記時間情報「Time」が「StandardTime」より大きい場合には表示に時間が掛かるためにその関連ファイルに係る深さレベルを大きくして、取得の優先順位を低くするのである。ここで、上記「StandardTime」は予め設定された標準的な時間であり、この時間内であれば略同一時間と見なすのである。

【0065】ステップS20で、上記巡回制御部5によって、上記ステップS18で得られた関連ファイル名と上記ステップS19で算出された「深さレベルFileLevel」とが対応付けられて、上記作業ワークメモリに設定された管理情報「巡回タイプ」に従って巡回ファイルリスト格納部7に格納される。すなわち、上記管理情報「巡回タイプ」が「タイプA」である場合には、巡回ファイルリスト71に関連ファイル名と深さレベルとを追加した後に、巡回ファイルリスト71を「深さレベルFileLevel」の昇順にソートしておく。これに対して、上記管理情報「巡回タイプ」が「タイプB」である場合には、巡回ファイルリスト71の最後尾に新たな関連ファイル名と深さレベルとを追加するのである。

【0066】ステップS21で、上記巡回制御部5によって、管理情報格納部8に格納されている制御パラメータ「巡回ページ数CurrentPageNum」の内容が、次のように更新される。すなわち、現在巡回処理中のファイルがトップファイルである場合には、次式によって巡回ページ数CurrentPageNumが更新される。

30 $CurrentPageNum = CurrentPageNum + 1$

また、上記関連ファイルである場合には、タグに基づいてタグレベル管理テーブル82の「ページフラグ」を参照し、次式によって巡回ページ数CurrentPageNumが更新される。

$CurrentPageNum = CurrentPageNum + \text{「ページフラグ」の値}$
ステップS22で、上記巡回制御部5によって、管理情報格納部8に格納されている制御パラメータ「巡回ファイル数CurrentFileNum」の内容が、次のように更新される。

40 $CurrentFileNum = CurrentFileNum + 1$

ステップS23で、上記巡回制御部5によって、管理情報格納部8に格納されている制御パラメータ「巡回容量CurrentVolume」の内容が、次のように更新される。

$CurrentVolume = CurrentVolume + \text{取得ファイルデータの容量}$

【0067】ステップS24で、上記巡回制御部5によって、巡回ファイルリスト71に関連ファイル名が在るか否かが判別される。その結果、在ればステップS26に進み、無ければステップS25に進む。ステップS25で、上記巡回制御部5によって、トップファイルリスト格納部

1 内に未処理のファイル名が存在するか否かを判別することによって、トップファイルリスト格納部 1 内の総てのトップファイルに対する巡回が終了したか否かが判別される。その結果、終了していなければ上記ステップ S 2 に戻って、次のトップファイルに対する処理に移行する。一方、終了していれば巡回制御処理動作を終了する。

【0068】ステップ S 26 で、上記巡回制御部 5 によって、上記作業ワークメモリに設定されている管理情報「巡回タイプ」の内容がチェックされる。ステップ S 27 で、上記巡回制御部 5 によって、上記ステップ S 26 においてチェックされた管理情報「巡回タイプ」の内容は「タイプ A」であるか否かが判別される。その結果、「タイプ A」であればステップ S 28 に進み、そうでなければステップ S 29 に進む。ステップ S 28 で、上記巡回制御部 5 によって、巡回ファイルリスト 7 1 の先頭から一つの関連ファイル名(例えば「ファイル B」)が読み出される。そうした後に、ステップ S 30 に進む。ステップ S 29 で、上記巡回制御部 5 によって、「巡回タイプ」は「タイプ B」であるから、巡回ファイルリスト 7 1 の最後尾から一つの関連ファイル名(例えば「ファイル B」)が読み出される。

【0069】ステップ S 30 で、上記巡回制御部 5 によって、上記ステップ S 28 あるいはステップ S 29 において読み出された関連ファイル名「ファイル B」が巡回ファイルリスト 7 1 から削除される。ステップ S 31 で、上記巡回制御部 5 によって、管理情報格納部 8 に格納されている制御パラメータ「深さレベル CurrentLevel」の内容が、上記ステップ S 28 あるいはステップ S 29 において読み出された関連ファイル名「ファイル B」に付加されている深さレベル FileLevel に更新される。

CurrentLevel=FileLevel

【0070】ステップ S 32 で、上記巡回制御部 5 によって、上記ステップ S 31 において更新された「深さレベル CurrentLevel」の内容と上記作業ワークメモリに格納されている「最大深さレベル MaxLevel」の内容とに基づいて、次式が成立するか否かが判別される。

最大深さレベル MaxLevel (= 3) < int(深さレベル CurrentLevel)

その結果、成立すればステップ S 24 に進み、成立しなければステップ S 33 に進む。ここで、int(x) は x に関する関数であり、x の小数点以下の切り捨てによる整数化関数である。

【0071】ステップ S 33 で、上記巡回制御部 5 によって、上記制御パラメータ「巡回ページ数 CurrentPageNum」の内容と上記作業ワークメモリに格納されている

「最大巡回ページ数 MaxPageNum」の内容とに基づいて、次式が成立するか否かが判別される。

最大巡回ページ数 MaxPageNum (= 100) < 巡回ページ数 CurrentPageNum

その結果、成立すればステップ S 24 に進み、成立しなければステップ S 34 に進む。

【0072】ステップ S 34 で、上記巡回制御部 5 によって、上記制御パラメータ「巡回ファイル数 CurrentFileNum」の内容と上記作業ワークメモリに格納されている「最大巡回ファイル数 MaxFileNum」の内容とに基づいて、次式が成立するか否かが判別される。

最大巡回ファイル数 MaxFileNum (= 500) < 巡回ファイル数 CurrentFileNum

その結果、成立すればステップ S 24 に進み、成立しなければステップ S 35 に進む。

【0073】ステップ S 35 で、上記巡回制御部 5 によって、ネットワークインターフェース部 3 に対して、「関連ファイル B」の取得が要求される。そうした後に、上記ステップ S 5 に戻って、ファイルキャッシュメモリ 2 内に「関連ファイル B」が存在するか否かの判別に移行する。そして、上記ステップ S 24 において巡回ファイルリスト 7 1 に関連ファイル名が無いと判別され、上記ステップ S 25 においてトップファイルリスト格納部 1 内の総てのトップファイルに対する巡回が終了したと判別されると巡回制御処理動作を終了する。

【0074】すなわち、本実施の形態においては、上記関連ファイル抽出手段を上記ステップ S 18 で構成し、上記深さレベル算出手段を上記ステップ S 19 で構成し、上記関連ファイル格納手段を上記ステップ S 20 で構成し、上記深さレベル監視手段を上記ステップ S 31, S 32 で構成し、上記ファイル削除手段及び削除指令手段を上記ステップ S 11, S 14 で構成し、上記メディアタイプ解釈手段を上記ステップ S 12 で構成し、上記ファイル容量監視手段を上記ステップ S 10, S 11 で構成するのである。

【0075】このように、本実施の形態においては、巡回すべきトップファイルのファイル名を格納するトップファイルリスト格納部 1 に、各ファイル名に関連付けて、ユーザによって設定された各トップファイルに関する最大深さレベル, 最大巡回ページ数, 最大巡回ファイル数のアクセス条件を格納しておく。そして、管理情報格納部 8 に格納されている制御パラメータ「深さレベル」, 「巡回ページ数」, 「巡回ファイル数」が上記最大深さレベル, 最大巡回ページ数, 最大巡回ファイル数に到達したら巡回処理を終了するようにしている。したがって、上記従来のハイパーメディア文書通信装置やソフトウェア「フリーロード」, 「波乗野郎」, 「インターネットマネージャ」のように、単に、関連度の高いページを選出したりページを分類したり取得可能ページ数を制限したりする場合よりも柔軟に且つ的確に、上記ネットワーク上におけるアクセス範囲を各トップファイル毎に制限して、制限されたキャッシュエリア内で必要なファイルを自動的にファイルキャッシュメモリ 2 に取得できるのである。

【0076】また、上記管理情報格納部 8 にはタグレベ

ル管理テーブル82を格納し、このタグレベル管理テーブル82には、2つのファイルが同一ページを構成するような関連状態であることを表す「IMGタグ」や、2つのファイルが別ページを構成するような関連状態であることを表す「ANCHORタグ」や、2つのファイルが同一ページを構成して且つ異なる時間情報を有するような関連状態であることを表す「METAタグ」等のタグに応じたタグレベルTAG_LEVELを登録している。そして、巡回制御部5は、上記トップファイルリスト格納部1に格納されたトップファイルの関連ファイルを取得する場合に、上記タグレベル管理テーブル82を参照して、ハイパーテキストのネットワーク上における現在の深さレベルCurrentLevelと取得しようとする関連ファイルのタグのタグレベルTAG_LEVELとに基づいて、上記取得しようとする関連ファイルの深さレベルFileLevelを算出するようにしている。したがって、本実施の形態によれば、上述のようにして算出された深さレベルFileLevelに基づいて、現在巡回処理が行われている上記ネットワーク上の深さレベルが最大深さレベルMaxLevelを越えないように監視することができる。

【0077】また、上記巡回制御部5によって、解析部6で抽出された関連ファイルの深さレベルFileLevelを算出する場合に、関連ファイルのタグがMETAタグである場合には、タグに記述された時間情報「Time」に応じて当該関連ファイルの深さレベルFileLevelの値を大きくして取得の優先順位を低くするようにしている。このように、上記時間情報「Time」を関連ファイルの深さレベルFileLevelに反映させることによって、表示に時間が掛かる関連ファイルの取得を後回しにして、重要な関連ファイルを優先して取得することができるのである。

【0078】また、本実施の形態においては、上記解析部6によって、巡回ファイルリスト71に登録された関連ファイル名に従って関連ファイルを取得する場合の巡回タイプとして、トップページから見て深さレベルの浅い順に横方向に巡回するタイプAと、トップページから順次深さレベルの深い方に巡回するタイプBとを定義する。そして、巡回制御部5は、トップファイルリスト格納部1に各ファイル名に関連付けて格納された管理情報「巡回タイプ」に基づいて、巡回ファイルリスト71に対する関連ファイル名の登録および読み出しを行うようにしている。したがって、ユーザによって指定された巡回タイプに従って、広く浅い関連ファイルの取得と狭く深い関連ファイルの取得との何れかの方法によって関連ファイルを取得できる。

【0079】また、本実施の形態においては、ファイルデータの種類を表すメディアタイプにそのメディアタイプに属するファイルデータを取得するか否かの取得フラグを対応付けたメディアタイプ管理テーブル83を上記管理情報格納部8に登録しておく。そして、巡回制御部

5は、上記メディアタイプ管理テーブル83を参照して、受け取ったファイルデータのメディアタイプを判断して取得するか否かを判定し、取得しない場合にはファイルキャッシュメモリ2に登録された当該ファイルのファイル名を削除するようにしている。したがって、上記メディアタイプ管理テーブル83を予め設定しておくことによって、関連ファイルとして取得するファイルデータを限定して、ユーザにとって有効なメディアタイプのファイルデータのみを取得できるのである。

10 【0080】また、本実施の形態においては、上記トップファイルリスト格納部1に、各ファイル名に関連付けて、ユーザによって設定された各トップファイルに関する最大巡回容量をアクセス条件として格納しておく。そして、巡回制御部5は、上記関連ファイルを受け取る毎に、当該関連ファイルを加味した取得ファイルの総容量が上記最大巡回容量を越えるか否かを監視する。そして、越える場合には、最後に受け取った関連ファイルのファイルデータの削除をネットワークインターフェース部3に指示するようにしている。したがって、特に携帯型の装置であってファイルキャッシュメモリ2の容量を多く取れない場合でも、上記容量内でユーザにとって有効なファイルを取得できるのである。

20 【0081】また、その場合に、取得ファイルの総容量が上記最大巡回容量を越えると直ちに巡回処理を停止するのではなく、最後に受け取ったファイルデータを削除して巡回処理を続行するので、次の巡回処理で受け取った関連ファイルのデータ容量が小さくて取得ファイルの総容量が上記最大巡回容量を越えない場合には、その関連ファイルは取得される。したがって、上記ファイルキャッシュメモリ2に格納し得る最大数のファイルを取得できるのである。

30 【0082】図11は、最大巡回容量MaxVolumeを「100kバイト」とし、巡回タイプとして「タイプB」を指定した場合のファイルの取得範囲を示す。この場合には、まず、ファイルAをメインファイルとしてイメージ1,2およびメディア1を含むトップページと、このトップページと同じ深さレベルのレベル「1」にある(リンクmetaで連結された)ファイルBが取得される。次に、縦巡回方式に従って、ファイルC、ファイルD、ファイルE、ファイルFと順次深さレベルの深い方に巡回処理が行われる。この時点での巡回容量CurrentVolumeは94kバイトであり、余裕は6kバイトである。したがって、次に、巡回制御部5は、レベル「2」の深さレベルに在るファイルGおよびファイルHのうち上記余裕内に入る(容量が6kバイト)ファイルHを取得する。その結果、上記巡回容量CurrentVolumeは、最大巡回容量MaxVolumeと同じ100kバイトとなり、図11においてハッチングで示す領域内に在るファイルが最大容量「100kバイト」以内で取得し得る最大数の有効なファイルとして取得される。このように、本実施の形態によれ

ば、ユーザによって上記アクセス条件(この場合には「最大巡回容量」と「巡回タイプB」)によって制限されたアクセス範囲に応じて、少ないキャッシュエリアでユーザが必要とするページを効率よく取得できるのである。

【0083】上述のように、本実施の形態によれば、ユーザインターフェース・ツール「ブラウザ」を用いた場合のように、通信中に対話によって指示しながらハイパーテキストを取得する必要が無く、次の指示を出すまでの間の通信の無駄や、何度も同じファイルにアクセスしてしまうロスを解消できる。さらに、上記従来の自動巡回機能を有するハイパーメディア文書通信装置や自動巡回機能を有するソフトウェアよりも柔軟に、巡回すべき上記ネットワーク上の範囲を制限して、少ないキャッシュエリアに必要なファイルを自動的に且つ的確に取得できるのである。

【0084】

【発明の効果】以上より明らかなように、請求項1に係る発明のハイパーテキスト自動取得装置は、ファイル取得制御部によって、ネットワークインターフェース部に対してトップファイル格納部あるいは関連ファイルリスト格納部に格納されたファイルの取得要求を出し、返送されたファイルデータを解析部に送出して関連ファイル名を抽出させ、抽出された関連ファイル名を上記関連ファイルリスト格納部に格納するファイル取得制御を、上記トップファイル毎にアクセス条件格納部に格納された最大深さレベルを含む上記ネットワークへのアクセス条件に従って行うので、各トップファイル毎に、柔軟且つ的確にアクセス範囲を制限して上記ネットワークへのアクセスを行うことができる。

【0085】特に、上記アクセス条件として「最大深さレベル」が含まれているので、上記ネットワーク上を進む際の深さレベルを、上記トップファイル毎に制限できる。すなわち、この発明によれば、上記ファイルキャッシュメモリの容量に応じて必要なファイルを的確に自動的に取得できるのである。

【0086】また、請求項2に係る発明のハイパーテキスト自動取得装置は、上記ファイルデータに含まれるタグとタグレベルとを対応付けてタグレベル管理テーブルに格納し、上記解析部は、関連ファイル抽出手段によって、上記タグが表す関連情報に基づいて関連ファイル名を抽出して上記タグと共に上記ファイル取得制御部に返送し、上記ファイル取得制御部は、深さレベル算出手段によって、上記タグレベルを用いて上記関連ファイルの深さレベルを算出し、関連ファイル格納手段によって関連ファイル名に付加して上記関連ファイルリスト格納部に格納し、深さレベル監視手段によって、上記関連ファイルの取得要求を出す際に当該関連ファイルの深さレベルが上記アクセス条件の最大深さレベルを越えないように監視するので、ファイル取得制御部による関連ファイ

ルの取得要求時における上記ネットワーク上のアクセス範囲を上記最大深さレベルを越えないように制限できる。

【0087】また、請求項3に係る発明のハイパーテキスト自動取得装置におけるタグは、2つのファイルが共に同一ページを構成するような関連状態を表すタグと、2つのファイルが異なるページを構成するような関連状態を表すタグを含んでいるので、上記両関連状態に対して異なるタグレベルを設定できる。

10 【0088】また、請求項4に係る発明のハイパーテキスト自動取得装置におけるタグは、2つのファイルが共に同一ページを構成し、且つ、異なる時間属性を有するような関連状態を表すタグを含み、上記ファイル取得制御部の深さレベル算出手段は、上記タグが有する時間情報を深さレベルの算出値に反映させるようになっているので、例えば、出力に時間が掛かるような関連ファイルの深さレベル値を大きくして、当該関連ファイルの取得の優先度を低めることができる。

20 【0089】また、請求項5に係る発明のハイパーテキスト自動取得装置におけるファイル取得制御部の関連ファイル格納手段は、上記アクセス条件格納部に第1のアクセス方法が格納されて場合には上記関連ファイルリスト格納部に関連ファイル名と深さレベルとを追加する毎に深さレベルの昇順にソートする一方、上記第2のアクセス方法が格納されている場合には上記関連ファイルリスト格納部の最後尾に追加し、上記ファイル取得制御部は、上記第1のアクセス方法の場合には上記関連ファイルリスト格納部の先頭から順に関連ファイル名と深さレベルとを読み出す一方、上記第2のアクセス方法の場合には、最後尾から順に読み出すので、上記第1のアクセス方法の場合には、上記トップファイルが含まれるトップページから見て上記ネットワーク上での深さが浅い順にアクセスが行われる一方、第2のアクセス方法の場合には、上記トップページから順次上記深さの深い方にアクセスが行われる。したがって、ハイパーテキストのネットワークからの広く浅い関連ファイルの取得と狭く深い関連ファイルの取得との何れかを、上記アクセス条件の1つとしてアクセス方法を指定することによって選択的に行うことができる。すなわち、この発明によれば、

40 上記アクセス方法をうまく指定することによって、限られたキャッシュエリア内で有効な関連ファイルを取得できる。

50 【0090】また、請求項6に係る発明のハイパーテキスト自動取得装置は、ファイルデータのメディアタイプとこのメディアタイプに属するファイルデータの取得情報とを対応付けて取得情報格納部に格納し、上記ファイル取得制御部は、受け取ったファイルデータのメディアタイプをメディアタイプ解釈手段によって解釈し、このメディアタイプの解釈結果に基づいて、削除指示手段によって上記取得情報格納部を参照して取得しないと判定

されたファイルデータの削除を上記ネットワークインターフェース部に指示するので、上記取得情報格納部に予め設定された取得情報に応じて、関連ファイルとして取得されるファイルデータのメディアタイプを限定できる。したがって、この発明によれば、ユーザが必要とするメディアタイプのファイルのみを取得できる。

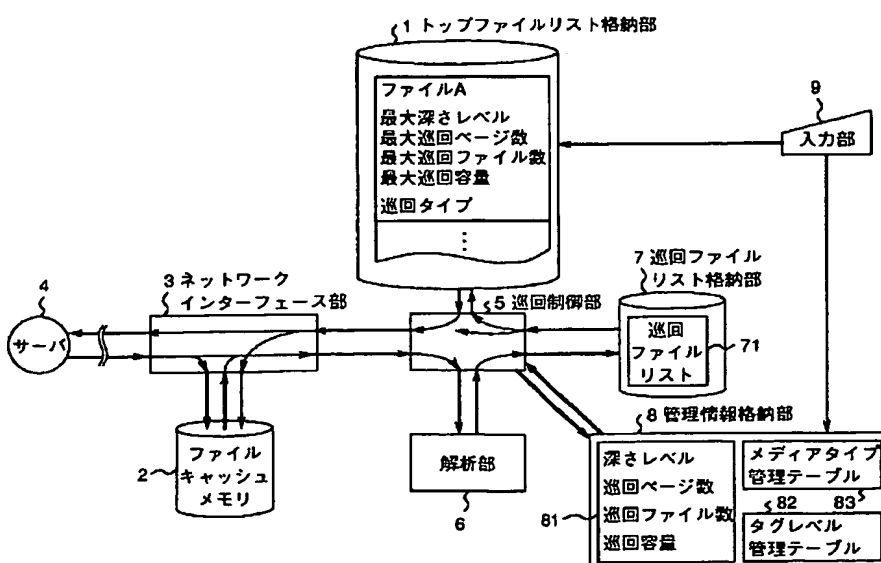
【0091】また、請求項7に係る発明のハイパーテキスト自動取得装置は、上記ファイル取得制御部はファイル容量監視手段を有して、関連ファイルを受け取る毎に、受け取った関連ファイルを加味した取得ファイルの総容量を監視し、上記アクセス条件としての最大取得容量を越える場合には最後に受け取ったファイルデータの削除を上記ネットワークインターフェース部に指示するので、特に携帯型の装置であって上記ファイルキャッシュメモリの容量が大きい場合でも、上記最大取得容量を最適に設定することによって、上記容量内で取得し得る最大数の有効なファイルデータを取得できる。

【0092】また、請求項8に係る発明のハイパーテキスト自動取得装置は、上記トップファイル格納部に格納されるファイル名、上記アクセス条件格納部に格納されるアクセス条件、上記タグレベル管理テーブルの要素、あるいは、上記取得情報格納部の要素は、入力部から入力・更新可能になっているので、ユーザは、上記入力部から上記トップファイル名、アクセス条件、上記両テーブルの要素を設定・更新して、ハイパーテキストのネットワーク上へのアクセス範囲を柔軟且つ的確に制限できる。

【図面の簡単な説明】

【図1】この発明のハイパーテキスト自動取得装置における機能ブロック図である。

【図1】



【図2】図1に示す機能を実現するハードウェアブロック図である。

【図3】図1に示す巡回ファイルリスト格納部に登録される巡回ファイルリストの一例を示す図である。

【図4】図1に示す管理情報格納部に格納されるタグレベル管理テーブルの一例を示す図である。

【図5】図1に示す管理情報格納部に格納されるメディアタイプ管理テーブルの一例を示す図である。

【図6】図1に示す巡回制御部、ネットワークインターフェース部および解析部によって行われる巡回制御処理動作のフローチャートである。

【図7】図6に続く巡回制御処理動作のフローチャートである。

【図8】図7に続く巡回制御処理動作のフローチャートである。

【図9】取得しようとするハイパーテキストの構造を示す図である。

【図10】深さレベルの説明図である。

【図11】最大巡回容量を100kバイトとした場合のファイルの取得範囲を示す図である。

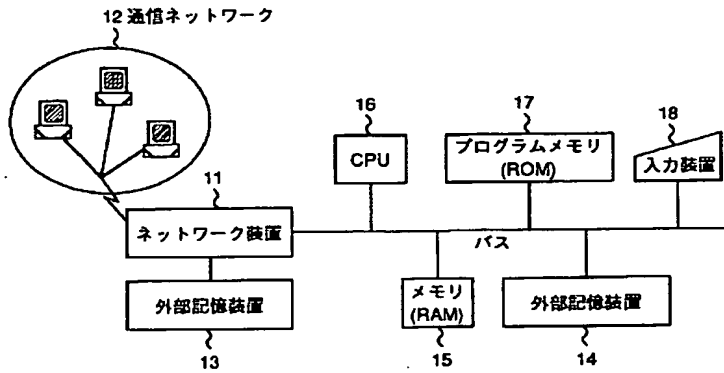
【符号の説明】

1…トップファイルリスト格納部、2…ファイルキャッシュメモリ、3…ネットワークインターフェース部、4…サーバ、5…巡回制御部、6…解析部、7…巡回ファイルリスト格納部、8…管理情報格納部、9…入力部、71…巡回ファイルリスト、81…制御パラメータ、82…タグレベル管理テーブル、83…メディアタイプ管理テーブル。

【図3】

ファイル名(URL)	深さレベル
http://server1/user/a.html	2
http://server1/user/c.gif	2
http://host2/xx/d.html	2.5
http://server1/user/b.html	3
http://host2/xx/	3
http://server3/yy.jpg	3
http://host2/xx/e.html	3

【図2】



【図5】

メディアタイプ	取得フラグ
text/html	取得する
text/*	取得する
img/*	取得する
application/*	取得しない
audio/basic	取得する
/	取得しない
...	...

【図4】

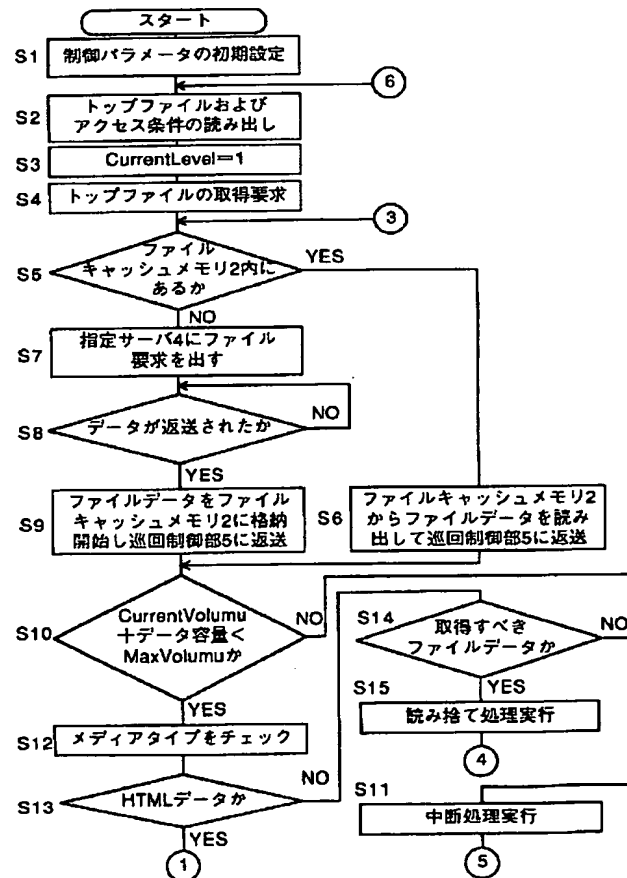
(a) 標準設定例

タグ種類	巡回フラグ	タグレベル	ページフラグ
IMG	1	0	0
ANCHOR	1	1	1
META	1	0.5	0
FRAME	1	0.5	0
...

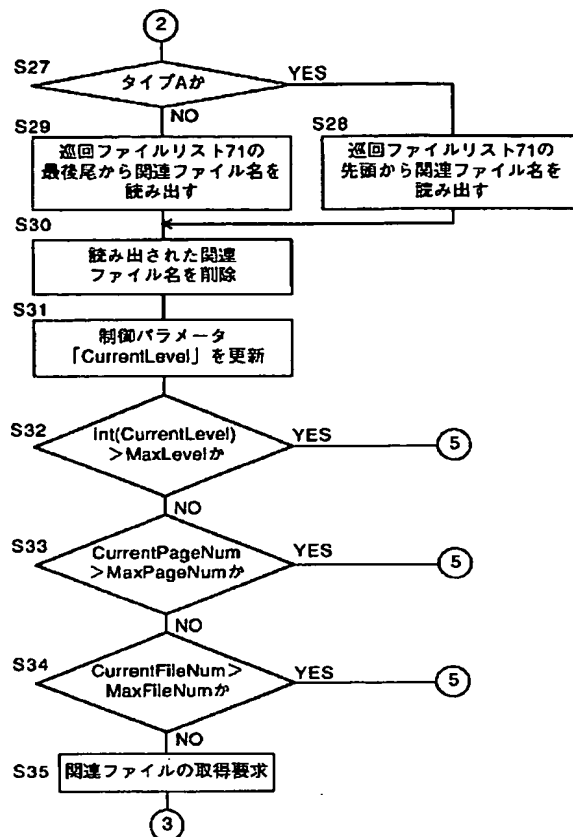
(b) ページ構成別による設定例

タグ種類	巡回フラグ	タグレベル	ページフラグ
IMG	1	0	0
ANCHOR	1	1	1
META	1	0	0
FRAME	1	0	0
...

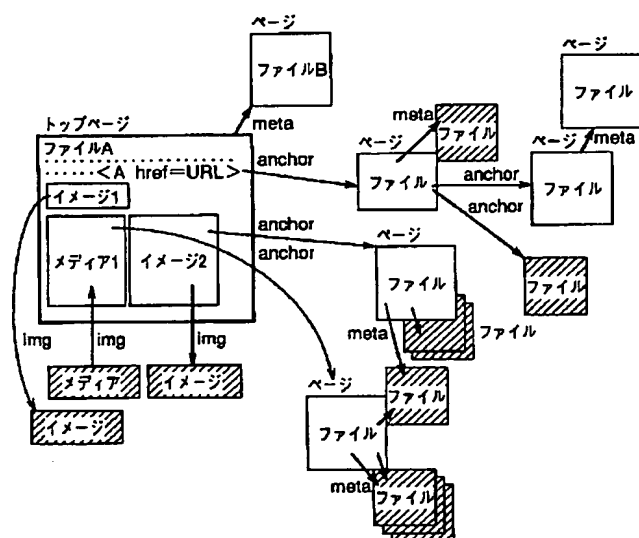
【図6】



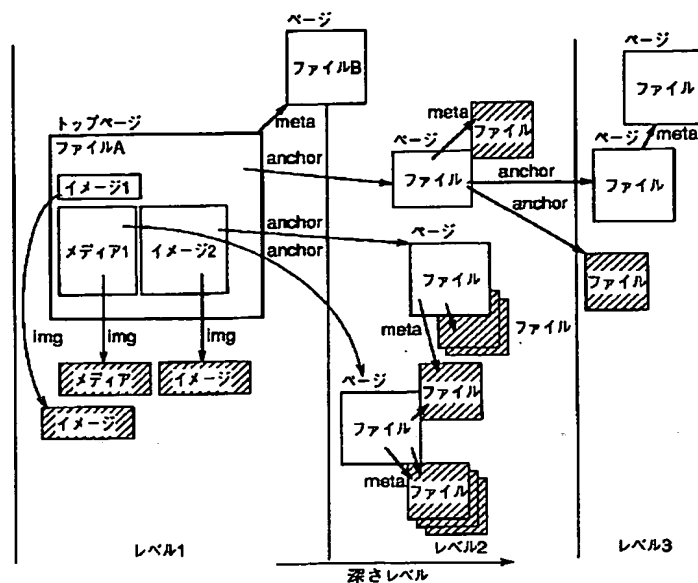
【図 8】



【图9】



【図 10】



【图 1 1】

